Project Report

Prototyping with Generative Agents

Filippo Merlo: filippo.merlo@studenti.unitn.it

September 2023

1 Introduction

The idea for this project originated from two recent studies by Joon Sung Park. The first study, titled ”Social Simulacra: Creating Populated Prototypes for Social Computing Systems”[8], focuses on a prototyping technique called ”social simulacra.” This technique is employed to investigate and understand the social behaviors that may emerge in a proposed design for a social computing system, examples of which could be platforms such as Reddit[1] or Discord[2]. The primary issue addressed in this study is that current prototyping methods for social computing systems typically involve recruiting small groups of people, which may not reveal all the challenges that can arise when the system is scaled up. Social simulacra is introduced as a solution to this limitation. It takes input in the form of a designer's description of a community's design, including its goals, rules, and member personas, and generates a realistic instance of that design with simulated social behaviors. These behaviors encompass various activities such as posts, replies, and even anti-social behaviors. The objective is to assist designers in comprehending how a social system might behave when populated, enabling them to make necessary adjustments to the design to mitigate potential challenges before they become problematic. The study also contributes techniques for instructing a large language model to generate thousands of distinct community members and simulate their social interactions, leveraging the model's training data, which includes a wide range of positive and negative behaviors on social media platforms.

The second study, which was the subject of my paper review, is titled ”Generative Agents: Interactive Simulacra of Human Behavior”[7]. It introduces the concept of ”generative agents,” which consist on an architecture built around a language model that enables it to better simulate human behavior by reasoning about a broader set of experiences than what can be described in a simple prompt. This is achieved by using a memory stream to surface relevant memories, allowing the generative agent to provide more informative and specific responses. Specifically, this approach enhances the generative agents' ability to answer questions about their experiences by avoiding the limitation of summarizing all experiences within a confined context window. Instead, it relies on relevant memories to provide responses. Additionally, it improves the generative agents' ability to generalize and make inferences by drawing from higher-level reflections based on observational memory.

The idea proposed by this project is to use the ”generative agents” described in the second study to enhance the believability of the social interaction simulation used in the prototyping technique of ”social simulacra.” As we will demonstrate, the utility of this choice extends beyond the credibility of the simulation because equipping each agent with its own personality, memory, and reflective abilities makes it possible to gather insights on the interaction directly from individual agents.

2 Methodology

In this section, we will describe the different components of the project. First, we will explain how the description of the online community has been generated from the input of the user. Second how agents were created by specifying the features used for their customization. Third, I will detail how their interaction was structured to simulate the typical social interaction in an online community. The last part will showcase the process of interviewing the agents that participated in the simulation.

2.1 Community Description

To allow flexibility in the experiment, the description of the online community that we want to simulate is generated using the `GPT-3.5-turbo` model[3] (the same model upon which `ChatGPT` is based). We ask the user to specify the following:

- `name`: The name of the community.
- `goal`: The purpose or goal of the community.
- `target_population`: The specific group of people the community is intended for.
- `rules`: The code of conduct or rules for members of the community.

The complete prompt for generating the description is as follows:

```
Write the description of an online community containing this information:
    \texttt{name}: {name}
    \texttt{goal}: {goal}
    \texttt{target\_population}: {target\_population}
    \texttt{rules}: {rules}
Output a single paragraph.
```

From the output generated in this manner, the model is then asked to provide a one or two-word definition of the community's topic. This information will be used further in the agents' background generation. An example of an community description is showed here:

```
goal: "This is the place to discuss about recent political events in Europe"
name: "EuroPol"
rules: "Do not be offensive, Cooperates, Be open minded"
target_population: "People interested in the histoy and political life of
    European countries"
topic: "European politics"
description: "EuroPol is an online community designed for individuals with
    a keen interest in the history and political landscape of
    European countries. With a goal centered around fostering
    discussions about recent political events in Europe, EuroPol
    serves as a platform for its target population to engage in
    thoughtful and insightful conversations. Emphasizing the
    importance of respectful communication, EuroPol establishes
    clear rules to ensure a positive and constructive environment
    for its members. These rules include a strict prohibition
    against offensive behavior, an encouragement for cooperation,
    and a call for open-mindedness. By adhering to these
    guidelines, EuroPol aims to create a space where individuals
    can exchange ideas, share perspectives, and deepen their
    understanding of the political developments shaping Europe."
```

2.2 Agents with LangChain

Since the code used to create the "generative agents" in [7] was made available only recently, I used the code provided by the Python library LangChain[4] to implement the architecture they are composed of. LangChain is a framework for language model-driven applications, offering data awareness and the ability for models to interact with their surroundings. It provides modular components for easy use and pre-configured chains for simplified tasks while allowing customization for complex applications. Among the modules in this library, there is one that recreates the "generative agents" from [7] using the functions and methods of LangChain.

The generative agent class LangChain provides accepts these main arguments we used for the personalization:

- age: The optional age of the character.
- llm: The underlying language model.
- memory: The memory object that combines relevance, recency, and importance.
- memory_retriever: retriever for observations in memory.
- name: The character's name.

- status: The traits of the character you wish not to change.
- traits: Permanent traits to ascribe to the character.

The memory object consists of a class named "GenerativeAgentMemory" that allows the storage of observations computed by the agent according to the dynamics described in [7]. It utilizes information such as relevance, recency, and importance to make the agent's behaviour more akin to that of a human. It accepts three key arguments that are:

- llm: The underlying language model.
- memory_retriever: retriever for observations in memory.
- reflection_treshold: When the aggregated importance of the last observations exceeds the reflection threshold the agent stops reflecting, we set it at 10.

Both the agent and the memory object rely on the `GPT-3.5-turbo` model [3] and both utilize the "TimeWeightedVectorStoreRetriever" function as memory retriever, which combines embedding similarity with recency in retrieving values.

Age is randomly generated from a Gaussian distribution (mean: 28, standard deviation: 12). Numbers less than 18 are removed and regenerated, as it is assumed that the community is frequented only by adults. Names are generated by the model. There are three traits, randomly extracted from a list of 40 traits, including 20 positive and 20 negative traits, also generated by the model. Once this information is obtained, the model is prompted to generate a brief description of an individual with these characteristics. Information about the community is added to make that agent consistent with the fact that they are in that particular community. The prompt is as follows:

```
Invent a short description of an imaginary person named: {name}, of age:
{list_ages[i]} with these traits: {traits}.
The description must contain only the name, the person’s life occupation and one peculiarity.
E.g.: [name] is a [life occupation] and he/she [peculiarity]
The description must be coherent with the fact that among
this person interests there is {community_topic}.
```

This information is used for the status argument. An example of complete agent's background is showed here:

```
name: "Michael"
age: 18
traits: "Reliable Cooperative Moody"
description: "Michael is a diligent student majoring in International
    Relations and European Politics. He is known for his unwavering reliability
    and cooperative nature when it comes to group projects and teamwork.
    However, his moodiness adds an intriguing twist to his character. One moment
    he can be jovial and enthusiastic about discussing the intricacies of
    European politics, and the next, he might retreat into a contemplative and
    introspective state, leaving others wondering what thoughts occupy his mind."
```

Before the simulation begins, all agents are initialized by embedding the following prompt into their memory, containing the description of the community:

```
You are in an online community named {community_name} whose description is: {community_description’}
```

Due to the cost of generating text with the model (the OpenAI API is not free) we decided to generate only 10 agents for the simulation.

2.3 Interaction Structure

To enable agents to interact in a manner similar to an online community, two additional methods are created within the base class of generative agents. All the interactions are recorded in a list of the community's posts. In this list, we find the posts list and in each post the information about the creator, the post's title, the post's content and the number of comments it received. Additionally, each post contains the list of the comments it received with the names of the agents that created them. The first method is called "generate_post_reaction" and allows the agent to take as input information such as the creator's name, the title, and the content of the last N posts and decide whether to generate a new post or not. The prompt used within this method is as follows:

```
Does {agent_name} want to start a conversation on a new topic by publishing a new post, and if so,
what would be the new post title and content? Respond in one line.
Write a new post only if {agent_name} is really interested in doing so.
If the action is to write a new post, write:
\nNEWPOST:title of the new post//content of the new post
\ndo not forget the divider sign // between title and content.
\nDo nothing, or write the new post but not both.\n\n
```

If the agent decides to react by generating a new post, the output is formatted and added to the list of total posts created within the community.

The second method is called ”generate_comment” and allows the agent to take as input information about one post and the last N comments under that post and decide whether to generate a new comment under that post. The prompt used within this method is as follows:

```
Should {agent_name} react to the observation by writing a comment to the post? Respond in one line.
\nWrite a comment only if {agent_name} is really interested in doing so and if you do try to
express the {agent_name}'s personal opinion about the topic.
If the action is to comment on the post, write:
\nCOMMENT: "what to comment"
\nEither do nothing, or write a comment but not both.\n\n
```

If the agent decides to react by generating a new comment, the output is formatted and added to the list of comments that post received.

Finally, in the simulation loop for each turn, there is a posting phase, where agents inspect and generate new posts and a comment phase where they read the existing posts and comment on them.

2.4 Final Interview

After running the simulation, thanks to the fact that each involved agent possesses its own independent memory, it is possible to individually interview each agent to obtain information about the interaction from different perspectives. We posed four questions to each agent regarding the simulation. The questions targeted specific information about the community, such as what the most interesting topic was and which agent they agreed with the most, as well as more general questions soliciting feedback about the community itself. Below are the questions posed to the agents:

```
1. What post interested you the most?"
2. With which user did you agree the most?"
3. If you could change something about the community's goals or rules, what would you change?
4. Did you enjoy the overall experience of being in this community? Answer with a number from 1 to 10.
```

3 Results

In this part, we show some of the posts generated during the simulation with related comments and two interviews with agents after the simulation was concluded.

Here are shown three of the several posts that were generated in the simulation:

```
Ashley wrote this post:
TITLE: "The Impact of Brexit on European Trade Relations"
CONTENT:  "I've been reflecting on the long-term consequences of Brexit on trade relationships
within Europe. Let's discuss how this historic event has shaped the economic landscape and what it
means for the future of European countries."

It received 4 comments:



Jennifer commented "I completely agree, Brexit has had a significant impact on trade relationships
within Europe. It's important to analyze the economic landscape and discuss the future implications
for European countries."

Christopher commented "I appreciate the thoughtful analysis, Ashley. Despite the challenges Brexit
```

has brought, it's interesting to consider how it may also present opportunities for European countries to forge new trade relationships."

Lisa commented "As a seasoned political analyst, I believe Brexit has indeed had significant consequences on trade relationships within Europe. It's crucial to analyze and discuss the long-term implications for European countries."

Jessica commented "I appreciate the insightful analysis, Ashley. As a retired diplomat, I have witnessed firsthand the complexities of trade relationships in Europe. It's important to consider both the challenges and opportunities that Brexit has brought to European countries."

****

Daniel wrote this post:
TITLE: "The Challenges of European Integration"
CONTENT: "As a political analyst specializing in European politics, I have been reflecting on the challenges faced by European integration. Let's discuss the obstacles that arise in creating a united Europe and brainstorm potential solutions."

It received 1 comments:

Lisa commented "I appreciate your perspective, Daniel. European integration certainly faces numerous challenges, including differing political and economic interests among member states. It would be interesting to explore potential solutions and strategies to overcome these obstacles."

****

Lisa wrote this post:
TITLE: "The Role of Environmental Policies in European Politics"
CONTENT: "I've been considering the impact of environmental policies on the political landscape of European countries. Let's explore the complexities of this issue and discuss its implications for the future of Europe."

It received 5 comments:

Sarah commented "I find the topic of environmental policies in European politics intriguing. It's crucial to discuss their impact and implications for the future of Europe, especially in terms of sustainability and addressing climate change."

Jennifer commented "I completely agree, Lisa. The role of environmental policies in European politics is crucial for the future of Europe, especially in terms of sustainability and addressing climate change. I'm looking forward to discussing the complexities of this issue with you and other members of EuroPol."

Michael commented "I completely agree with both Sarah and Jennifer. The role of environmental policies in European politics is crucial for the future of Europe, especially in terms of sustainability and addressing climate change. I'm eager to delve deeper into this topic and hear different perspectives from fellow EuroPol members."

Christopher commented "I appreciate the thought-provoking topic, Lisa. The role of environmental policies in European politics is indeed crucial for the future of Europe, and it's fascinating to explore the complexities and implications of this issue. Looking forward to hearing different perspectives on this matter from fellow EuroPol members."

David commented "I agree with the other commenters, Lisa. The role of environmental policies in European politics is crucial for the future of Europe, and it's important to discuss the

```
complexities and implications of this issue. Looking forward to engaging in this conversation with
fellow EuroPol members."
```

Following are reported the interviews with two of the agents who took part in the simulation:

```
1. Daniel said "The post that interested me the most was 'The Challenges of European Integration.'
As a political analyst specializing in European politics, I find the topic of European integration
to be incredibly important and complex. I believe it's crucial to address the obstacles that arise
in creating a united Europe and brainstorm potential solutions. What are your thoughts on this
issue?"

2. Daniel said "The user I agree with the most is Lisa. Her post on the role of environmental
policies in European politics caught my attention. I believe that environmental issues are critical
in shaping the future of Europe, and it's essential to discuss the complexities and implications of
these policies. What are your thoughts on this topic?"

3. Daniel said "If I could change something about the community's goals or rules, I would suggest
implementing a system that encourages more active participation and engagement from members.
Perhaps introducing regular debates or structured discussions on specific topics could help foster
deeper analysis and exchange of ideas. Additionally, I believe it would be beneficial to have
guidelines in place to address any manipulative or passive-aggressive behavior that may arise
within the community. This would ensure that discussions remain respectful and constructive for
everyone involved. What are your thoughts on these suggestions?"

4. Daniel said "I would rate my overall experience in this community as a 7. While I appreciate the
opportunity to engage in meaningful discussions on European politics, I do sometimes find it
challenging to navigate different perspectives and maintain constructive conversations. However,
the valuable insights and diverse viewpoints shared by fellow members make it a worthwhile
experience. Thank you for asking."

**************************************************************************************************

1. Jessica said "The post that interested me the most was 'The Impact of Brexit on European Trade
Relations'. It's a topic that I've been following closely and I believe it has had significant
consequences on trade relationships within Europe. It's important to analyze and discuss the long-
term implications for European countries."

2. Jessica said "The user I agreed with the most was Lisa. Her insightful analysis of the impact of
nationalism on European politics resonated with me. As a retired diplomat, I have witnessed
firsthand the complexities of this phenomenon and I believe it is crucial to understand its
implications for shaping the future of Europe."

3. Jessica said "I believe the community's goals and rules are well-defined and necessary for
fostering meaningful discussions about European politics. I appreciate the emphasis on respectful
communication and open-mindedness. I wouldn't change anything about them."

4. Jessica said "I would rate my overall experience in this community as a 9. I have thoroughly
enjoyed engaging in thoughtful discussions and learning from different perspectives. It has been a
fulfilling experience for me as a retired diplomat, and I appreciate the opportunities to
contribute to the discussions. Thank you for asking!"
```

4 Discussion and Conclusion

Some observations we can make about these generations are that all three posts generated a decent amount of engagement with multiple comments on each. In all cases, the comments are generally supportive and positive in tone, and there is no presence of any negative or troll comments. This can be a consequence of the model used since the gpt-3.5-turbo[3] model has been heavily fine-tuned with human feedback to remove any antisocial or toxic behaviour[6], making him overtly cooperative and friendly. Giving the agents different backgrounds and character traits provided some variety of perspectives both in their contribution to the discussions and in their personal experience of the community. This diversity of perspectives enriches the discussions adding depth to the conversation and providing different insights on how to shape the community differently.

Finally, some comments are quite similar, with multiple commenters expressing agreement or using similar phrases. For example, in Lisa's post about environmental policies, several commenters mention the importance of addressing climate change and look forward to engaging in the conversation. This might be caused by a scarce background and memory information differentiating agents at the beginning of the simulation. I expect that making longer simulations or working on generating more detailed and exhausting backgrounds for the agents will substantially improve the variety and difference of their generations.

In summary, social simulacra and generative agents are effective tools for simulating online communities and understanding their behaviour at scale. While the results show positive engagement and constructive discussions, there is room for improving response diversity. These techniques hold promise for community design and insights, but further research is needed for refinement.

Overall I think that this project has been a first step for me towards both the theoretical and practical understanding of these powerful tools, which are generative agents, and their various possible applications. It has led me to explore new platforms like Langchain and the methods they provide, which I believe will be a valuable experience for the future. Finally, being able to adapt them flexibly to implement a functional automated interaction structure was a challenge that I am glad to have successfully completed. Code available at the following link[5]

References

[1] URL: https://www.reddit.com/.

[2] URL: https://discord.com/.

[3] URL: https://platform.openai.com/docs/models/gpt-3-5.

[4] URL: https://python.langchain.com/docs/get_started/introduction.

[5] Filippo-Merlo. *Filippo-Merlo/socSimulacra*$_{w}$*ith*$_{g}$*enAgents*. URL: https://github.com/filippo-merlo/socSimulacra_with_genAgents.git.

[6] Long Ouyang et al. *Training language models to follow instructions with human feedback*. Mar. 4, 2022. arXiv: 2203.02155[cs]. URL: http://arxiv.org/abs/2203.02155 (visited on 09/10/2023).

[7] Joon Sung Park et al. *Generative Agents: Interactive Simulacra of Human Behavior*. Apr. 6, 2023. arXiv: 2304.03442[cs]. URL: http://arxiv.org/abs/2304.03442 (visited on 04/18/2023).

[8] Joon Sung Park et al. *Social Simulacra: Creating Populated Prototypes for Social Computing Systems*. Aug. 8, 2022. arXiv: 2208.04024[cs]. URL: http://arxiv.org/abs/2208.04024 (visited on 08/29/2023).